

TIME-RESOLVED CARDIAC CT RECONSTRUCTION USING THE ENSEMBLE KALMAN FILTER

A K George

M D Butala, R A Frazin, F Kamalabadi, and Y Bresler

NHLBI
National Institutes of Health
Bethesda, Maryland 20892

Coordinated Science Laboratory and Dept of ECE
University Of Illinois At Urbana-Champaign
Urbana, Illinois 61801

ABSTRACT

We propose an algorithm to solve the problem of Time-Resolved Cardiac Computed Tomography (CT). The algorithm reconstructs a snapshot of the moving heart at any time instant from CT projection data acquired over a single heart-cycle. The object is modeled by a spatio-temporal state-space model, and an ensemble Kalman Filter (a Monte-Carlo approximation to the Kalman filter) is used to assimilate the sequentially acquired projection data. Simulation results of the dynamic NCAT cardiac phantom, under the fan-beam geometry and a two-source CT system, show reconstructions that are free of the motion artifacts that mar conventional methods.

Index Terms— Cardiac, Dynamic Tomography, Time-resolved, Computed Tomography, Kalman filter

1. INTRODUCTION

Cardiac disease is the leading cause of death in the industrialized world. Recent interest [1] in reconstructing high-resolution images of the moving heart is driven both by the need to improve existing clinical applications, such as the detection of heart-wall motion abnormalities or the determination of parameters such as peak ejection fraction, and the potential for future applications such as the study of the relationship of cardiac motion to perfusion, stenosis, ischemia or infarction.

Recognizing the potential of cardiac CT (in providing superior spatial resolution and much shorter imaging times than MRI, for example) equipment manufacturers have introduced products such as the 320-slice Toshiba scanner that allows for coverage of the whole heart with a circular trajectory and the dual-source Siemens scanner that allows for the imaging of the quiescent heart phase within a single heart cycle.

The majority of early work on cardiac CT imaging has been on how best to combine data from multiple heart-cycles to reconstruct the data in the quiescent heart phase (when the heart is almost static). The actual reconstruction algorithms used are variations of conventional backprojection, sometimes modified by a weighting function to reduce motion artifacts. There has been some preliminary work on reconstructing an image of the heart in the nonquiescent phase by framing it as an optimization problem that maximizes data consistency (matching the measured projections to the projections of the reconstructed object) while making reasonable assumptions about nature of the object such as the continuity or smoothness of the objects and its borders in space and time or the smoothness of the motion-vector field ([1, 2] and the references therein).

This research was partly supported by the Research Board at the University of Illinois at Urbana-Champaign.

We present an algorithm based on the ensemble Kalman filter [3] (EnKF) : a Monte Carlo approximation to the Kalman filter, that has previously been used to address problems in weather prediction, acoustic tomography and remote sensing. The algorithm in this paper is the *localized* EnKF[4] that has been adapted to cardiac CT. We use the quiescent-phase reconstruction as a prior and choose the Kalman noise model based on our knowledge of the motion of human organs. The results of our numerical experiment with the NCAT (NURBS [nonuniform rational B-spline] based cardiac torso phantom) [5] cardiac phantom display reconstructions that are free of the motion artifacts that mar conventional reconstruction methods. This promising method can be extended to other medical imaging modalities such as MRI.

2. PROBLEM FORMULATION

2.1. The dynamic CT problem

Our goal is to reconstruct a movie of a beating heart from fan-beam projections. While in this paper we describe the reconstruction of a single transverse slice of the heart, the algorithm can be generalized to 3D volumes. The fan-beam projections are obtained in a *two-source* CT system in which a source of x-rays revolves around the object in a circle of radius D . As shown in Figure

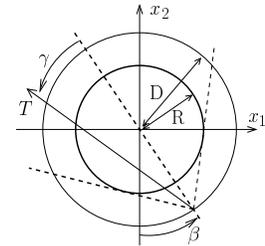


Fig. 1. The single-source fan-beam CT geometry with a source-radius D , source angle β and fan angle γ .

1, each fan-beam projection $p_{\beta}(\gamma, t)$ parameterized by the source angle β , fan-angle γ and time t , consists of a set of line-integrals of the object $f(\vec{x}, t)$. Since the sources in our two-source system are separated by an angle of $\pi/2$ rad, at every time instant t_i exactly two fan-beam projections are measured: with the first x-ray source at source angle β_i and the second source at source angle $\beta_i + \pi/2$. The discrete set of measurements made at the time-instant t_i is

$$\{p_{\beta_i}(\gamma_1, t_i), p_{\beta_i}(\gamma_2, t_i), \dots, p_{\beta_i}(\gamma_F, t_i), \\ p_{\beta_i + \pi/2}(\gamma_1, t_i), p_{\beta_i + \pi/2}(\gamma_2, t_i), \dots, p_{\beta_i + \pi/2}(\gamma_F, t_i)\} \quad (1)$$

The sources rotate at three revolutions per second with each

source generating $P(\approx 1000)$ projections every revolution. For typical heart rates of 60 to 100 beats/min the length of a single heart cycle is typically between 1.0 s and 0.6 s corresponding to 2 to 3 CT gantry rotations. From projections gathered over as little as a single heart cycle we wish to reconstruct the snapshots $f(\vec{x}, t_i)$ of the moving heart at any time instant within that measurement time frame.

An image of a static object can be reconstructed from fan-beam projections with a range of $\pi + \Gamma$ rad using a filtered backprojection algorithm, where Γ is maximum extent of the fan-angle ($\Gamma = 2 \arcsin(R/D)$ in Figure 1). The difficulty in applying reconstruction methods that are designed for static objects is that there can be significant movement of the heart during the time frame in which $\pi + \Gamma$ rad of projections are gathered. Consequently the projection measurements do not correspond to a single static object and are inconsistent. Therefore, applying the filtered backprojection algorithm to the dynamically gathered projections results in motion artifacts in the reconstructed snapshots.

2.2. Linear MMSE estimation

We set up the problem in a state-space formulation in which we estimate the state x_i of the object at time-index i by measurements made at various other time-instants. We choose the state of the object to be the N^2 -length vector of pixel values of the $N \times N$ image. The dynamic evolution of the object, and the measurement of the dynamically evolving object, by fan-beam projections, is modeled as follows :

$$\vec{x}_{i+1} = F_i \vec{x}_i + \vec{u}_i \quad \vec{y}_i = H_i \vec{x}_i + \vec{v}_i \quad (2)$$

where $\vec{x}_i, \vec{u}_i \in \mathbb{R}^{N^2 \times 1}$ and $\vec{y}_i, \vec{v}_i \in \mathbb{R}^{M \times 1}$.

The first equation is the time-update equation, which describes what happens to the state vector \vec{x}_i as time progresses. The matrix F_i describes how the state at a time index $i + 1$ is related to the state at the previous time index. The vector \vec{u}_i , which we call the *state noise*, is a random vector whose covariance matrix is $E[\vec{u}_i \vec{u}_i^T] \triangleq Q_i$. Inclusion of the state noise in our model allows for error in the model of the state evolution. (For convenience, from now on we drop the vector notation $\vec{\cdot}$ for state and measurement vectors.)

The second equation is the measurement equation, which describes how y_i , the measurement made at time index i , is related to the state vector x_i at that time instant. The elements of the length- $2F$ vector y_i belong to the set described in (1). In particular $y_i[m] = \bar{y}_i[m] + v_i[m]$ where

$$\bar{y}_i[m] = \begin{cases} p_{\beta_i}(\gamma_m, t_i) & \text{if } m = 1, 2, \dots, F \\ p_{\beta_i + \pi/2}(\gamma_{m-F}, t_i) & \text{if } m = F + 1, F + 2, \dots, 2F \end{cases} \quad (3)$$

The matrix H_i describes tomographic projection in the fan-beam geometry and $\bar{y}_i = H_i x_i$. The m^{th} row of H_i , denoted $h_{i,m}$, performs a weighted summation of the pixels of the image, i.e., the elements of the vector x_i , to produce the line integral $\bar{y}_i[m]$. So $\bar{y}_i[m] = h_{i,m} x_i$. The vector v_i is a random vector, representing the *measurement noise* with covariance matrix $E[v_i v_i^T] \triangleq R_i$ is a diagonal matrix derived from the standard CT signal dependant additive noise approximation to the Poisson noise model.

The problem is to estimate the state of the object at all time indices $\{x_i : i = 1, 2, \dots, T\}$, given the measurements $\{y_i : i = 1, 2, \dots, T\}$ at all those time indices. The Kalman filter finds the linear minimum mean-square error (MMSE) estimate of x_i given all the previous measurement. It does so in the following recursive

manner. At the beginning of the algorithm, a guess of the state of the object (i.e., $\hat{x}_{1|0} \triangleq x_0$) at time $i = 1$, and a guess of the state of the error covariance matrix of that guess (i.e., $P_{1|0} \triangleq \Pi_0$) is made. Then the Kalman filter recurses through every time index ($i = 1, 2, \dots, T$) and estimates $\hat{x}_{i|i}$, the state of the object at that time index based on past measurements. At every time index two updates are made: (a) the measurement update and (b) the time update:

$$\begin{aligned} \text{(a) Measurement update : } & \hat{x}_{i|i} = \hat{x}_{i|i-1} + K_i(y_i - H_i \hat{x}_{i|i-1}) \\ & \text{where } K_i = P_{i|i-1} H_i^T (H_i P_{i|i-1} H_i^T + R_i)^{-1} \\ & \text{and } P_{i|i} = P_{i|i-1} - K_i H_i P_{i|i-1} \end{aligned}$$

$$\begin{aligned} \text{(b) Time update : } & \hat{x}_{i+1|i} = F_i \hat{x}_{i|i} \\ & P_{i+1|i} = F_i P_{i|i} F_i^T + Q_i \end{aligned}$$

In the time update step the state vector $\hat{x}_{i+1|i}$ and the covariance matrix $P_{i+1|i}$ are estimated on the basis of the forward model and the past measurements $\{y_{i'} : i' = 1, 2, \dots, i\}$. In the measurement update step the current measurement y_i is incorporated by altering the estimate of the state vector by the difference $[y_i - H_i \hat{x}_{i|i-1}]$ scaled by the Kalman gain matrix K_i .

2.3. Motion model and state noise

Cardiac motion can be described as motion and deformation of objects (heart chambers and vessels, and their walls) and fluids (blood, with or without contrast agent). This motion is modeled as the linear interpolation of the pixels from frame to frame as follows. Suppose the 2D interpolant used is $\Upsilon(\vec{r})$ (where \vec{r} is the continuous Cartesian spatial variable); then the underlying continuous image $x_i^c(\vec{r})$ is expressed in terms of the pixels of the image $x_i(n)$ as: $x_i^c(\vec{r}) = \sum_{n'} x_i(n') \Upsilon[\vec{r}' - \vec{r}(n')]$. Here $\vec{r}(n)$ is the spatial coordinate associated with the n^{th} pixel. The motion of the object between time-frames is then modeled as $x_{i+1}(n) = x_i^c(\vec{r}(n) + \vec{u}_n)$ where \vec{u}_n is the motion of the n^{th} pixel in frame $i + 1$.

The true motion \vec{u}_n is *a priori* unknown. As explained later in this section, an estimate \vec{d}_n of \vec{u}_n is made before the Kalman filter is applied. The true motion $\vec{u}_n = \vec{d}_n + \vec{\eta}_n$, where $\vec{\eta}_n$ is a 2D random vector with a pdf Φ that describes the uncertainty in the *a priori* motion estimate. The time interval between consecutive projections is so small that \vec{u}_n and, therefore, $\vec{\eta}_n$ are on the order of a pixel length. The choice of Φ governs the covariance Q_i of the state noise \vec{u}_i in (2) and, in turn, the structure of the Kalman gain matrices K_i .

The expected value of the pixel in the $(i + 1)^{\text{th}}$ frame depends on the probability density function Φ : $E[x_{i+1}(n)] =$

$$E_{\Phi}[x_i^c(\vec{r}(n) + \vec{d}_n + \vec{\eta}_n)] = \int x_i^c(\vec{r}(n) + \vec{d}_n + \vec{\eta}_n) \Phi(\vec{\eta}_n) d\vec{\eta}_n \quad (4)$$

Similarly, the variance of the pixel value is

$$\sigma^2[x_{i+1}(n)] = E_{\Phi} \left\{ \left[x_i^c(\vec{r}(n) + \vec{d}_n + \vec{\eta}_n) - E[x_{i+1}(n)] \right]^2 \right\} \quad (5)$$

which can be approximated in the discrete-domain by a weighted sum over pixels in its neighborhood:

$$\sigma^p[x_{i+1}(n)] = \sum_{n' \neq n} w[\vec{r}(n) - \vec{r}(n')] |E[x_{i+1}(n)] - x_{i+1}(n')|^p$$

Here w determines the weights and $p > 0$ determines the power by which the differences between the pixel and its neighbors are scaled.

While a direct application of the trapezoidal integration rule would lead to $p = 2$ in the above equation, our numerical experiments suggest that using $p = 1$ produces reconstructions with better edge preservation.

Consider $\sigma^2[x_{i+1}(n)]$ in different regions of the image. In flat regions of the image, where it is constant on a disc of radius R_Φ around the point $\vec{r}(n) + \vec{d}_n$, the expected value $E_\Phi[\hat{x}_{i+1|i}(n)] = x_{i|i}^c(\vec{r}(n) + \vec{d}_n)$ and the variance $\sigma^2[x_{i+1}(n)]$ in (5) is 0. In contrast, in regions that are not flat (such as points along the boundaries of objects) the variance is high and increases as the range of pixel-values (in a radius- R_Φ -disc) increases.

The probability density function Φ that we use is such that $E_\Phi[x_{i+1}(n)] \approx x_i^c(\vec{n} + \vec{d}_n)$. This expression can be expressed as a linear operation $E_\Phi[\vec{x}_{i+1}] = F_i \vec{x}_i$ where F_i is a sparse matrix that depends on Υ , the 2D interpolant used in the image model. Additionally, we make the assumption that the state noise is uncorrelated between pixels so that the covariance matrix Q_i of the state noise is a diagonal matrix with $Q_i(n, n) = \sigma^2[x_{i+1}(n)]$ as defined in (5).

A consequence of our choice of state-error noise model is that the Kalman gain matrix tends to make greater alterations to the pixels of the image that are in a neighborhood with a large pixel-value range (such as pixels near edges of objects rather than those in flat regions).

In scenarios such as ours where the time interval between reconstructed frames is small, the identity map $F_i(\vec{x}) = \vec{x}$ might suffice. Alternatively, an estimate of the motion is obtained from the projection data by reconstructing low-pass estimates of the image frames (naively) from low-passed fan-beam projections and then using an Optical Flow algorithm [6] to estimate the motion field \vec{d}_n between frames.

2.4. Initialization

During the *diastasis* phase of the heart cycle, the heart is almost static. Projections gathered during this *quiescent* period are used to estimate the prior x_0 of the EnKF. Using the method of Parker's weights it is possible to reconstruct x_0 from a source-angle range of $\pi + \Gamma$. In our two source system, it is necessary for each source to move by $\pi/2 + \Gamma$ rad for the two sources to provide the adequate range of source angles $\pi + \Gamma$. At 3 revolutions per second, this is achieved in less than 0.12 s (assuming source-radius $D = 57$ cm and radius of the thorax $R = 20$ cm, and therefore $\Gamma = 0.72$ rad). For typical heart-rates the quiescent period is longer than 0.12 s.

3. REDUCTION OF COMPUTATIONAL COST

3.1. The ensemble Kalman filter

For large N the Kalman filter becomes expensive. For example, the measurement update steps cost $O(N^4 M)$ operations at every step of the iteration, and the $N^2 \times N^2$ covariance matrix $P_{i|i'}$ needs to be stored and updated. The EnKF uses statistical (Monte Carlo) techniques to estimate the evolving dynamic states and their covariance matrices. Unlike the standard Kalman filter, the EnKF maintains not one, but an ensemble of L estimated states $\{x_{i|l}^l : l = 1, 2, \dots, L\}$. The EnKF estimates the relevant covariance matrices by averaging over this ensemble.

To initialize, we use the initial guess x_0 and the initial covariance estimate Π_0 to generate, using a pseudorandom noise generator, an ensemble of L initial states: $x_{1|0}^l \sim \mathcal{N}(x_0, \Pi_0)$. The random variables $y_i^l \sim \mathcal{N}(y_i, R_i)$ and $u_i^l \sim \mathcal{N}(0, Q_i)$ in the EnKF are gen-

erated using standard pseudorandom number generators.

(a) Measurement update: $\hat{x}_{i|i}^l = \hat{x}_{i|i-1}^l + K_i(y_i^l - H_i \hat{x}_{i|i-1}^l)$

$$\hat{K}_i = (\hat{P}_{i|i-1} H_i^T) [H_i (\hat{P}_{i|i-1} H_i^T) + R_i]^{-1}$$

$$\text{where } \hat{P}_{i|i-1} H_i^T = \frac{1}{L-1} \sum_{l=1}^L \hat{x}_{i|i-1}^l (H_i \hat{x}_{i|i-1}^l)^T$$

$$\bar{x}_{i|i-1} = \frac{1}{L} \sum_{l=1}^L x_{i|i-1}^l \text{ and } \tilde{x}_{i|i-1}^l = x_{i|i-1}^l - \bar{x}_{i|i-1}$$

(b) Time update : $\hat{x}_{i+1|i}^l = F_i(\hat{x}_{i|i}^l) + u_i^l$

3.2. Further optimizations

Sequential measurement update: If the noise of the individual measurements is independent as in the case of CT, then the covariance matrix R_i is diagonal, and the measurement update step can be performed *sequentially*. At every time index, instead of processing all M measurements (i.e., detectors) in one step, the update can be performed in M steps. This means that the relevant measurement matrices H are not $M \times N^2$ but are instead $1 \times N^2$ in size, and the computation of \hat{K}_i in the meas. update step is reduced to the inversion of a scalar rather than a matrix. Notice that the $N^2 \times N^2$ covariance matrix \hat{P} is never explicitly computed. It suffices, instead, to compute the $N^2 \times 1$ vector $\hat{P}_{i|i-1} h_{i,m}^T$.

Localization: For reasons of computational practicality we expect the dimensions of the state vector \vec{x} ($= N^2$) to be, typically, much larger than the size of the ensemble L . While the estimates of the EnKF approach those of the canonical Kalman filter for large ensembles, the behavior of the EnKF is not guaranteed to be similar to that of Kalman filter for small ensembles. Localization is a tool, related to other statistical methods known as covariance shrinkage or tapering, that is used to improve the behavior of the EnKF for small ensembles. Localization additionally provide reductions in computational cost. Because of localization the estimated matrix $\hat{P}_{i|i}(h_{i,m})^T$ (an $N^2 \times 1$ matrix) is nonzero only on pixels that are within a fixed correlation radius (c_r) of the line along which the line-integral is performed.

Reduction to ROI: The computational cost of the EnKF algorithm scales with the number of pixels to be reconstructed. The dynamic region of interest (ROI), the cardiac region, is a small fraction of the whole torso but since x-rays are transmitted and absorbed by the noncardiac objects the noncardiac objects have to be accounted for in the reconstruction. This is done by estimating the non-ROI region from the quiescent phase, subtracting the projections of this non-ROI region from the full dynamic projections, modeling the uncertainty in the estimate as measurement noise in (2) and finally smoothly patching the ROI and non-ROI regions together if necessary.

3.3. Computational cost

With the above optimizations it can be shown that the cost of each time-index of the EnKF is $O(N^2 L c_r^2)$ where L is the size of the ensemble and c_r is the correlation radius of the localization mask. In comparison the conventional Kalman filter which incorporates a sequential measurement update and sparse covariance matrices $P_{i|i-1}$ and $P_{i|i}$ (with c_r^2 nonzero entries per row) can be shown to cost $O(N^2 M c_r^2) = O(N^3 c_r^2)$ computations per time-index. In practice the ensemble size L is a fraction of N (in our numerical experiments $L \approx N/2$), and therefore, both the EnKF and the Kalman filter display the same *order* of computational complexity $O(N^3 c_r^2)$. Our

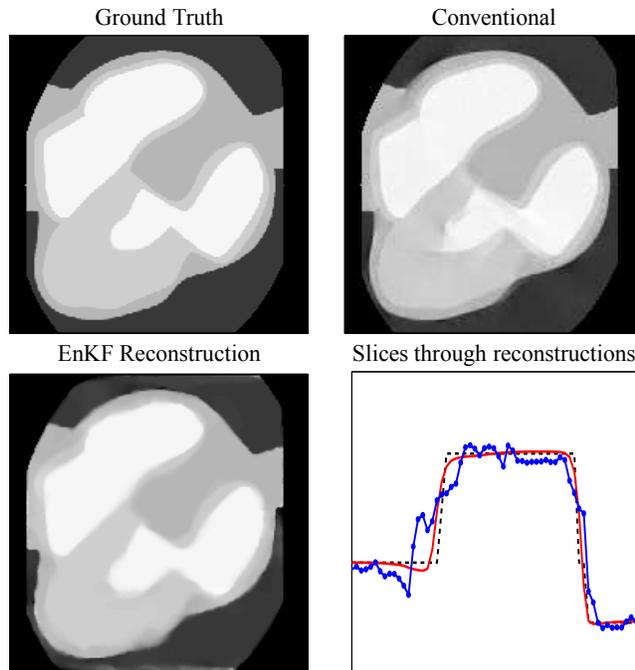


Fig. 2. A single time-frame of the NCAT phantom. Also shown is a detail of an oblique slice through each: ground truth (dashed line), conventional (dotted line) and EnKF (solid line).

numerical experiments suggest that the EnKF is more efficient in practice than the Kalman filter.

It may also be noted that the EnKF framework lends itself to a parallelized implementation where the members of the ensemble are processed in parallel. In addition, nonlinear prior motion models F are easily incorporated into the EnKF. We incorporate a median-filtering (on a 3×3 pixel window) into every fifth frame of the motion model as it is seen to improve the visual quality of the reconstruction.

4. NUMERICAL EXPERIMENTS

We test our method on the dynamic NCAT phantom[5] of the human torso with a (high) heart rate of 100 beats/min. At every time instant, two fan-beam projections (separated by $\pi/2$ rad in source angle) are obtained and the sources take 1200 source-positions per revolution at a speed of 3 revolutions per second. The dynamic ROI is 190×190 pixels. Projections are gathered for about the length of a single heart beat (0.6 sec) which is equivalent to $1200 \times 3 \times 0.6 = 2160$ time instants t_i . To save on computations the EnKF is run on only half of this data (data from every alternate time-instant is discarded).

Figure 2 shows a single frame of the phantom and of the reconstructions from projections using the conventional method and our method. It is clear that the reconstructed frame using our method is free of the motion artifacts (such as unclear and misaligned object boundaries) that mar the conventional method. A detail of a profile of the three images along an oblique cut is shown in the bottom right subplot of Figure 2. The misalignment of the edge of the conventional reconstruction is better seen in the profile.

In Figure 3 the oblique cut from Figure 2 is displayed for the full heart cycle. The EnKF reconstruction is seen to match the dynamic behavior of the phantom, while the conventional reconstruction displays motion artifacts.

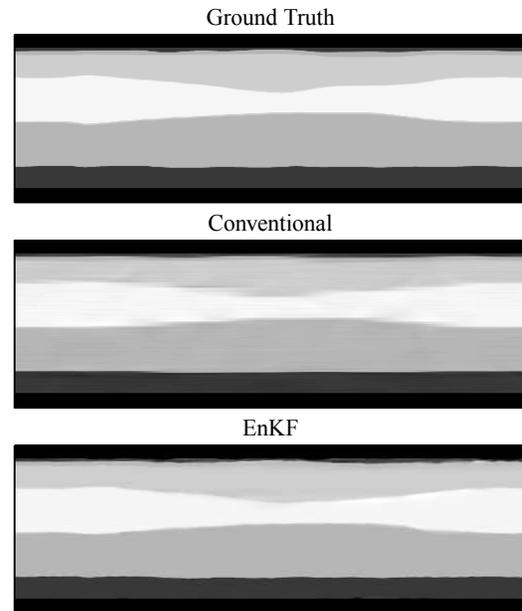


Fig. 3. Slices through the reconstructed image over a full heart cycle.

5. CONCLUSION

We have demonstrated the application of the EnKF to cardiac CT reconstruction. Our numerical simulations show that reconstructions of the heart during nonquiescent cardiac phases are free of the motion artifacts present in conventional reconstructions. There is potential for future work in improving the quality of reconstructions by fine-tuning the state-noise model, adapting the EnKF to other dynamic imaging scenarios, such as the case of single-source CT, and other modalities such as MRI. There is also room for theoretical work in characterizing the conditions for the optimal performance of the EnKF.

6. REFERENCES

- [1] K. Taguchi, Z. Sun, W.P. Segars, E.K. Fishman, and B.M.W. Tsui, "Image-domain motion compensated time resolved 4D cardiac CT," *Proceedings of SPIE*, vol. 6510, pp. 651016, 2007.
- [2] S. Bonnet, A. Koenig, S. Roux, P. Hugonnard, R. Guillemaud, and P. Grangeat, "Dynamic X-ray Computed Tomography," *Proceedings of the IEEE*, vol. 91, no. 10, pp. 1574–1587, 2003.
- [3] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, 2003.
- [4] M.D. Butala, R.A. Frazin, Y. Chen, and F. Kamalabadi, "A Monte Carlo technique for large-scale dynamic tomography," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007, vol. 3, pp. 1217–1220.
- [5] W.P. Segars, *Development and Application of the New Dynamic NURBS-based Cardiac-torso (NCAT) Phantom*, Ph.D. thesis, University of North Carolina at Chapel Hill, 2001.
- [6] T. Gautama and MA Van Hulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1127–1136, 2002.